

## 2. HIV の遺伝的多様性とバイオフィーマティクス

山口 由美

産業技術総合研究所 生物情報解析研究センター 統合データベース解析チーム

HIV-1 の表面タンパク質である gp120は、ウイルスの標的細胞への侵入において重要な役割がある一方、タンパク質の変異が免疫系による認識からの逃避に関わっている。つまり、アミノ酸の変化には生存に不利になる（負の淘汰）場合と有利（正の淘汰）になる場合の両方がある。このことは、アミノ酸変化の起こる場所、どんなアミノ酸に変わるか、によって事情が違うはずであり、各アミノ酸サイトの機能や立体構造上の位置と密接な関係にあると考えられる。筆者らは、HIV-1 gp120の各アミノ酸サイトの多型の再評価、分子進化進化機構、および適応進化に関する解析研究を行っており、特に以下の2つ問題に着目した。1) 保守的な進化あるいは適応的な進化、2) HIV-1 の株間による標的細胞の違いに関連している gp120のアミノ酸サイトはどこか。利用出来る塩基配列データを用いてアミノ酸サイトごとの詳細な解析を行い、生物学的情報やタンパク質の立体構造の情報との対応を調べた。得られたデータや結果は、進化的な理解に分子生物学的根拠を与えるだけでなく、機能予測につながることもある。

### HIV-1 gp120のアミノ酸の多様性の再評価

HIV の進化速度は非常に速く、系統間での高い遺伝的多様性がある。HIV の変異を解析、実態を理解することは、ワクチンなど治療法の開発、進化や起源の理解の両方にとって重要であり、塩基配列データも蓄積し続けている。データの蓄積は、HIV のタンパク質の機能解析のような生物学的データについても同様であり、タンパク質の立体構造データも利用出来るようになった。遺伝子の塩基配列やアミノ酸配列の多様性や進化のパターンをサイトごとに解析すると、変異導入実験などの生物学的知見との対応が取れるようになる（図1）。

ウイルスの表面タンパク質は、細胞への侵入における重要な機能をもつので、ウイルスの増殖能を維持する上でアミノ酸配列の多様性には機能的な制約がかかっている。その一方で、表面タンパク質のアミノ酸の変化は抗原性を多様にし、免疫系からの逃避に役立つ場合もある。つまり、アミノ酸の変化には生存に不利になる場合と有利になる場

合の両方がある。このことは、アミノ酸の変化の起こる場所、どんなアミノ酸に変わるか、によって事情が違うはずであり、各アミノ酸サイトの機能や立体構造上の位置と密接な関係にあると考えられる。レセプター結合に関わるアミノ酸サイトは保存的なのだろうか？そして抗原性の変化に関わるアミノ酸サイトは多様性に富むのだろうか？

HIV-1 gp120の配列が決定されて間もなく、配列の保存性の程度から、可変的なV領域と保存性の高いC領域が同定された<sup>2,3)</sup>。V3ループの抗原性や細胞指向性における重要性の知見もあって、V領域を中心に研究が進む傾向にあった。1998年にgp120のコアの部分の立体構造が発表された<sup>3)</sup>。これは幾つかのV領域を部分的に欠くものであったが、gp120の全体像がおおよそ分かるようになり、コアの部分の各アミノ酸残基の立体構造上の位置のデータが利用可能になった。また、配列データの蓄積も著しく、沢山の配列データを用いることによって、サイトごとの多型の特徴を良く分析できるようになった。筆者らは、HIV-1のサブタイプBのgp120の利用出来る配列をデータベースより収集し<sup>4)</sup>、多型を分析した。

アミノ酸の多様性の解析方法には様々なやり方があり、研究者や目的によって異なるだろうが、筆者らは、以下の3通りの方法で評価した<sup>5)</sup>（図2）。

1) 出現するアミノ酸の種類の数。受け入れられるアミノ酸の種類範囲。機能的制約の程度の指標となる。（得

### 連絡先

〒135-0064 東京都江東区青海2-41-6  
TEL: 03-5531-8550 (代表) 5531-1566 (直通)  
FAX: 03-5531-8551  
E-mail: yyamaguc@jbirc.aist.go.jp

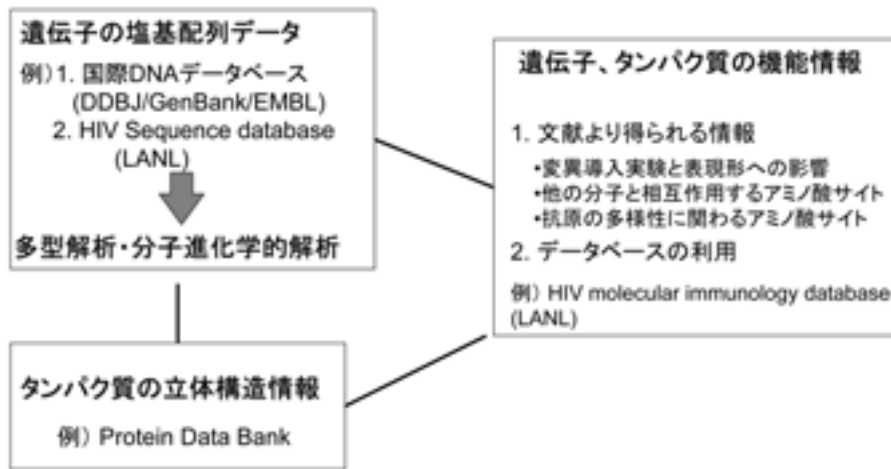


図1 ウイルスの遺伝子多型解析における情報の活用

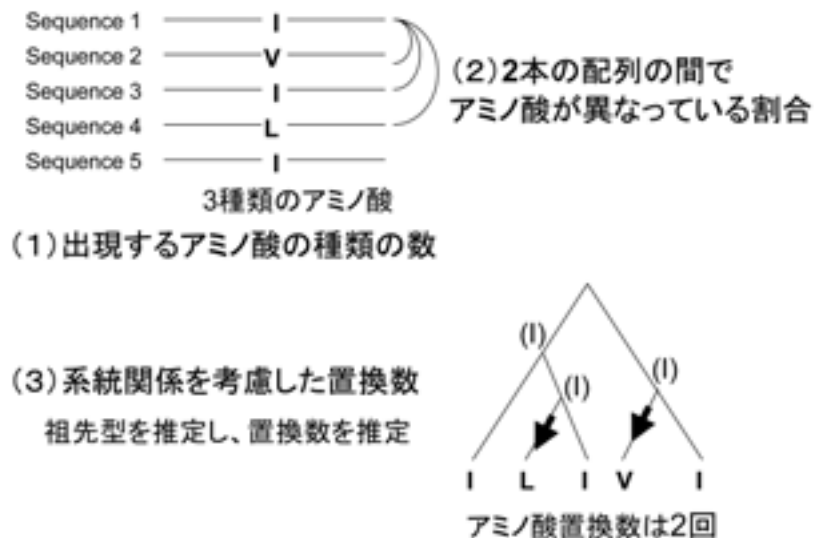


図2 アミノ酸サイトごとの多様性の評価  
文献5による。

られる値は、解析した配列の数に依存する)

- 2) 2本の配列の間でアミノ酸が異なっている割合。これは、集団遺伝学における、平均ヘテロ接合頻度と同じ考えである。集団内での変異の量の指標。各サイトにおけるアミノ酸の出現頻度を用いて計算できる。
- 3) 系統関係を考慮したアミノ酸置換数。変わりやすさの指標。分子系統樹を作成して、祖先の状態を推定し、置換数を推定する。(得られる値は、解析した配列の数に依存する)

gp120の多様性の再評価によって、アミノ酸の多様性の程度はC領域、V領域、共にサイト間で大きく異なっていること、C領域においても非常に高い多様性を示すサイトが存在することが示された。特に、タンパク質の外側に

あるアミノ酸残基では、側鎖の向きが内側か外側かによって、アミノ酸の多様性の程度が大きく異なっている。C3領域の $\alpha$ ヘリックスやC4領域の $\beta$ シートでは、多様なサイトと保存的なサイトの並びに周期性がある。

gp120のCD4やケモカインレセプターとの結合に関わるアミノ酸サイトは保存的だろうか？これは機能的な重要性より保存的であると予想される。しかし一方で、免疫系の認識のターゲットにもなり得る。gp120のサイトごとのアミノ酸多型をレセプターとの結合に関わるサイトについてまとめたところ<sup>5)</sup>、大部分のサイトが保存的であった。しかし、非常に多様性に富むサイトも含まれていた。実際、CD4結合部位の周辺にはエピトープの存在が知られているので、多様性の高いサイトの存在も説明できる。また、

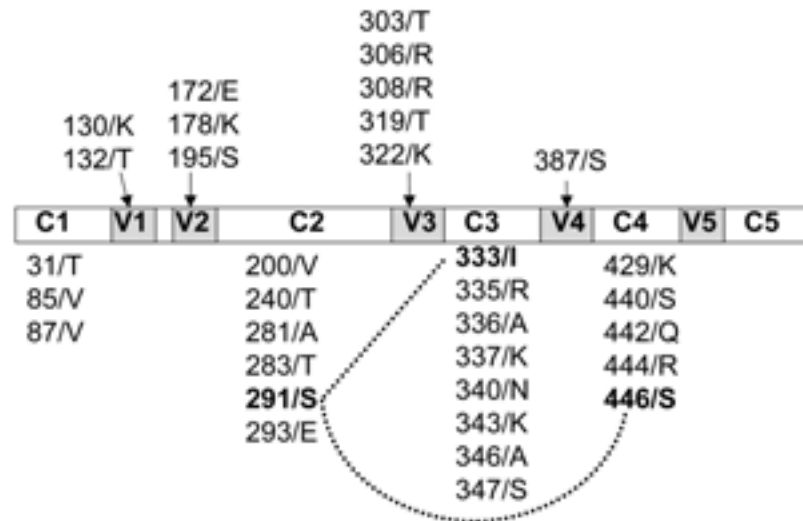


図3 HIV-1 gp120 (サブタイプB) で正の淘汰がかかっていると推測されるアミノ酸サイト

文献5による。サイトの番号は、HXB2株を基準にしている。  
点線は、立体構造上で近い関係にあることを示す。

CD4結合に関わるサイトの中には、側鎖でなく主鎖が結合に関わっているものもあり、側鎖が変化しても結合に重大な影響が出ない場合もありえる。

#### 保守的な進化それとも適応的な進化か？

前述したように、ウイルスの表面タンパク質のアミノ酸の変化には生存に不利になる場合（負の淘汰）と有利になる場合（正の淘汰）の両方がある。各アミノ酸サイトにおける、正と負の淘汰の相対的な強さをみるために、同義置換と非同義置換の置換数の推定と比較を行った。この推論の根拠は、分子進化の中立説での予測に基づく。中立説での予測では、遺伝子の非同義置換の速度は、同義置換速度よりも低くなり、非同義置換の速度が抑えられる程度は、タンパク質の機能的制約の程度で説明できる。様々な生物種の殆どのタンパク質において、非同義置換の速度は同義置換速度よりも抑えられており、中立説をサポートする根拠となっている。しかし、例外的かもしれないが、寄生性病原体のタンパク質や、免疫系に関する分子などでは、非同義置換の速度の上昇がみられる<sup>6)</sup>。もし、非同義置換の速度が同義置換速度を統計的に有意に超える場合は、中立仮説は棄却される。何らかの自然淘汰が働いて、アミノ酸置換が積極的に起こっている可能性が高い。HIVやインフルエンザウイルスのように、比較できる配列の数が十分に多く多型の程度も高い場合は、このような解析を各コドンポジションで詳細に分析することが可能である<sup>7)</sup>。筆者らは、HIV-1 gp120 (サブタイプB) 全長の配列を集め、系統関係を考慮して、各コドンポジションにおける同義置換と非同義置換の置換数を推定と比較を行った。

HIV-1 gp120では、非同義置換数が有意に高く、正の淘汰が主に働いていると推測されるアミノ酸サイトが33箇所検出された<sup>5)</sup> (図3)。タンパク質の立体構造との対応を調べたところ、この33箇所のアミノ酸サイトでは、殆どの場合、アミノ酸残基の側鎖がタンパク質の外側に突き出ていることが分かった。これらのサイトについて、抗原性や表現形に関する効果が報告されているかどうかを、文献からの情報と HIV molecular immunology database<sup>8)</sup>を用いて確認した。これらのサイトの中には、配列上では離れていても、立体構造上では近くに集まるものもあり、不連続エプトープを形成している可能性がある。この部位は、糖鎖付加部位の存在から“silent face”と理解されていた<sup>9)</sup>。しかし、分子進化的な解析結果は、速すぎるアミノ酸置換が起こっているサイトを検出し、エプトープの存在を示唆している。

HIV-1 gp120での正の淘汰、負の淘汰をサイトごとに検出する解析研究は、文献8, 10にも報告があり、検出されたサイトは、用いた配列セットや解析手法によって、若干異なっている。また、同様な解析はインフルエンザのHAタンパク質でも詳細に行われており、立体構造との対応が興味深い。

#### 細胞指向性の違いに関わるアミノ酸サイトはどこか？

HIV-1の細胞指向性は株によって違いがあり、マクロファージ指向性を持つか、T細胞指向性かを持つかは、第2レセプターであるケモカインレセプターの使用の違いによって説明されている<sup>11,12)</sup>。細胞指向性の決定基については、gp120のV3領域についての知見が豊富であるが、他

```

Phenotype1_strain1 ACDEFGHKLMPQRSTVWY
Phenotype1_strain2 ACDDFGHKIMNPQKSTVWY
Phenotype1_strain3 ACDEFGHKLMPQRSTVWY
Phenotype2_strain1 VCDEFGHKLMPQRSIVWY
Phenotype2_strain2 ACDEFGHKLMPQRSTVWY
Phenotype2_strain3 ACDEFGHKLMPQRSTLWY

```

グループ間の2本の配列の間でアミノ酸が異なっている割合

図4 表現形の違いに関わるアミノ酸サイトを見出すための解析例  
文献14による。

表1 HIV-1 サブタイプBの細胞指向性と gp120の配列の特徴  
文献14による。

セカンドレセプター使用	V3ループのアミノ酸配列の塩基性の程度	440番目 (C4領域) のアミノ酸
CCR5 (R5)	低い	塩基性アミノ酸 Arg, Lys
CCR5とCXCR4 (R5X4) CXCR4 (X4)	高い	非塩基性で体積が小さめ Gly, Ser, Glu, Thr, Gln

の領域の関与についても報告がある。筆者らは、細胞指向性の情報のある gp120の配列を集めて、細胞指向性の特徴づけるアミノ酸多型を同定するための解析を試みた。

HIV-1株 (サブタイプB) の第2レセプター使用の情報を、Domsらのデータベース<sup>13)</sup>より入手した。マクロファージ型のHIVは主にCCR5を使う(R5型)のに対し、T細胞型のHIVは主にCXCR4を使う。両方のレセプターを使う株は、R5X4型と分類されている。筆者らは、X4の表現形があるかどうかに関心があったので、HIVの21の株を2つのグループに分けて解析を行った<sup>14)</sup>。まず、分子系統樹を作成して、表現形の変化が進化の上で独立に何度も起こっていることを確認した。2つのグループ間のアミノ酸の多型の程度を解析し(図4)、配列をランダムにリサンプルすることによるブートストラップ法を用いて統計的な検定を行った。

その結果、3つのアミノ酸サイトが表現形とリンクしたのものとして検出された。そのうち2つはV3領域の11番目と25番目のサイトであり、変異と細胞指向性との関係が良く研究されているものであった。C4領域の440番目の多型は、非常にはっきりした特徴(表1)を示した。R5型では塩基性アミノ酸で、X4型では非塩基性で体積の小さめのアミノ酸であり、電荷はV3の特徴と逆で相補的な関係になっている。V3と440番目のサイトは立体構造上近い位置にある。構造上の安定性のために、相補的な進化をしているのだろうか? 440番目のサイトのコドンの使用頻度を調べたところ(文献5で用いた186本の配列)、非塩基性のアミノ酸(X4型、R5X4型に見られる)が使用しているコドンは、リジン、アルギニン(R5型に見られる)の使用されているコドンと殆どの場合一文字違いであることが分かった。おそらく、塩基性アミノ酸と非塩基性アミ

ノ酸との間での変化が1塩基置換によって、頻繁に起こっていると考えられる。Hoffmanら<sup>15)</sup>も同様な目的でgp120の配列を解析した。配列をV3領域のアミノ酸配列から予測された表現形でグループ分けした点は、直接的ではないが、多数の配列を解析に用いることができ、細胞指向性に関与するアミノ酸サイトの候補を指摘した。

ここで紹介した表現形と多型の関連解析では、アミノ酸置換のみを対象にしていたが、それ以外の多型である挿入、欠失、また糖鎖結合部位などの独自の着眼点で分析するようにも発展できる。

## おわりに

本稿で紹介した研究は、分子進化の原動力や適応進化についての興味から行われたものである。遺伝子配列の変異や進化のパターンを構造や機能についての情報と照らし合わせることは、大変貴重な機会であった。このようなアプローチは、総合的な理解を助けるだけでなく、相互の情報間の新しい関係を見出せる可能性も秘めていると期待している。更なる機能解析のための変異導入のデザインなど、実験支援のインフラに発展する方向性もある。遺伝子多型研究において、病原性ウイルスの場合は治療法対策などのため、データ量や分析の程度において、他の生物種よりも研究が先行している。分子生物学において、「構造、機能、進化の三位一体」を望む声を聞いたことがある。HIV研究は、既に三位一体の強さでもって取り組まれているのだと思う。

## 謝 辞

ここで紹介した研究は、国立遺伝学研究所の五條堀孝教授、京都大学ウイルス研究所の速水正憲教授、三浦智行助

教授との共同研究であり、ここに深く感謝する。

## 文 献

- 1) Starcich BR, Hahn BH, Shaw GM, McNeely PD, Modrow S, Wolf H, Parks ES, Parks WP, Josephs SF, Gallo RC, et al. : Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45** : 637-648, 1986.
- 2) Modrow S, Hahn BH, Shaw GM, Gallo RC, Wong-Staal F, Wolf H. : Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates : prediction of antigenic epitopes in conserved and variable regions. *J. Virol.* **61** : 570-578, 1987.
- 3) Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. : Structure of an HIV gp120 envelope glycoprotein in complex with the CD 4 receptor and a neutralizing human antibody. *Nature* **393** : 648-659, 1998.
- 4) Korber B, Hahn B, Foley B, Mellors JW, Leitner T, Myers G, McCutchan F, and Kuiken C (ed.). : Human retroviruses and AIDS 1997 : a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N. M., 1997.
- 5) Yamaguchi-Kabata Y, Gojobori T. : Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. *J. Virol.* **74** : 4335-4350, 2000.
- 6) Endo T, Ikeo K, and Gojobori T. : Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* **13** : 685-690, 1996.
- 7) Suzuki Y, Gojobori T. : A method for detecting positive selection at single amino acid sites. *Mol Biol Evol.* **16** : 1315-1328, 1999.
- 8) Korber B, Brander C, Haynes B, Koup R, Moore J, and Walker B (ed.). : HIV molecular immunology database 1998. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N. M., 1999.
- 9) Wyatt R, Sodroski J. : The HIV-1 envelope glycoproteins : fusogens, antigens, and immunogens. *Science.* **280** : 1884-1888, 1998.
- 10) Yang, Z. 2001. Maximum likelihood analysis of adaptive evolution in HIV-1 gp120 env gene. Pacific Symposium on BioComputing 2001. 226-237, 2001.
- 11) Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, Cayanan C, Maddon PJ, Koup RA, Moore JP, Paxton WA. : HIV-1 entry into CD 4 + cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381** : 667-673, 1996.
- 12) Feng Y, Broder CC, Kennedy PE, Berger EA. : HIV-1 entry cofactor : functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272** : 872-877, 1996.
- 13) Doms RW, Edinger AL, Moore JP : "Coreceptor use by primate lentiviruses" In HIV Sequence Database (<http://hiv-web.lanl.gov/>), 1998.
- 14) Yamaguchi-Kabata Y, Yamashita M, Ohkura S, Hayami M, Miura T. : Linkage of amino acid variation and evolution of human immunodeficiency virus type 1 gp120 envelope glycoprotein (subtype B) with usage of the second receptor. *J Mol Evol.* **58** : 333-340, 2004.
- 15) Hoffman NG, Seillier-Moisewitsch F, Ahn J, Walker JM, Swanstrom R. : Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V 3 loop. *J Virol.* **76** : 3852-3864, 2002.

# Sequence variation of HIV and bioinformatics

**Yumi Yamaguchi-Kabata**

Integrated Database Team, Biological Information Research Center, National Institute of  
Advanced Industrial Science and Technology (AIST)  
2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan  
E-mail : yyamaguc@jbirc.aist.go.jp

The envelope glycoprotein of human immunodeficiency virus type 1 (HIV-1) interacts with receptors on the target cell and mediates virus entry by fusing the viral and cell membranes. To maintain the viral infectivity, amino acids that interact with receptors are expected to be more conserved than the other sites on the protein surface. In contrast to the functional constraint of amino acids for the receptor binding, some amino acid changes in this protein may produce antigenic variations that enable the virus to escape from recognition of the host immune system. Therefore, both positive selection (higher fitness) and negative selection (lower fitness) against amino acid changes are taking place during evolution of surface proteins of parasites. To elucidate the evolutionary mechanisms of the whole HIV-1 gp120 envelope glycoprotein at the single site level, we collected and analyzed all available sequence data for the protein. By analyzing 186 sequences of the HIV-1 gp120 (subtype B), we reevaluated amino acid variability at the single site level, and estimated the numbers of synonymous and nonsynonymous substitutions at each codon position to detect positive and negative selection. We identified 33 amino acid positions which may be under positive selection. Some of these positions may form discontinuous epitopes. We also analyzed amino acid sequences to find amino acid positions responsible for usage of the second receptor. We found that, in addition to the V3 loop, amino acid variation at residue 440 in C4 region is clearly linked with the usage of CXCR4.